

Spring 4-2014

Red Light Green Light: A Novel Approach to Studying the Interaction between Enhancers and Gene Promoters

Jordan E. Vellky

Follow this and additional works at: https://ecommons.udayton.edu/uhp_theses



Part of the [Biology Commons](#)

eCommons Citation

Vellky, Jordan E., "Red Light Green Light: A Novel Approach to Studying the Interaction between Enhancers and Gene Promoters" (2014). *Honors Theses*. 7.

https://ecommons.udayton.edu/uhp_theses/7

This Honors Thesis is brought to you for free and open access by the University Honors Program at eCommons. It has been accepted for inclusion in Honors Theses by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.

Red Light Green Light: A Novel Approach to Studying the Interaction between Enhancers and Gene Promoters



Honors Thesis

Jordan E. Vellky

Department: Biology

Advisor: Thomas Willams, Ph.D.

April 2014

Red Light Green Light: A Novel Approach to Studying the Interaction between Enhancers and Gene Promoters

Honors Thesis

Jordan E. Vellky

Department: Biology

Advisor: Thomas Williams, PhD

April 2014

Abstract

The human genome consists of over 3 billion base pairs, but only about 2% of this extensive supply of genetic information is recognized as a functional protein coding sequence. The remaining 98% was once considered to be “junk DNA” that lacked functional elements. Recently, this assumption has been replaced by an understanding that the non-coding genome contains many functional elements involved in gene regulation. These elements include promoters, or the region where gene expression is initiated, and enhancers, which communicate to promoters information about the cells in a body and when a specific gene’s expression should be ON or OFF. My thesis research aims to develop and utilize a transgenic system to track the communication between enhancers and promoters in the fruit fly species *Drosophila melanogaster*, and eventually map the functional sequences for each gene of interest. Because promoters and enhancers are genetic components of all eukaryotic organisms, the system developed here can be applied to the genes of other organisms, including humans.

Acknowledgements

I would like to thank my advisor, Dr. Thomas Williams, Eric Camino, The University of Dayton Biology Department, The University of Dayton Honors Program, and the rest of the Williams’ Lab for supporting me throughout my thesis research



Table of Contents

Abstract	Title Page
Chapter 1: Introduction	2
Chapter 2: Materials and Methods	12
Chapter 3: Results	25
Chapter 4: Discussion	33
Chapter 5: Appendices	39
Chapter 6: References	54

CHAPTER 1: INTRODUCTION

Throughout the course of evolution, one consistency that pervades the growing complexity between organisms is the increasing size of the genome. Prokaryotic organisms, and even single-celled eukaryotes, possess smaller genomes than more complex eukaryotes, like humans (Figure 1). Multicellular animal (metazoan) genomes are large in gene number (>10,000) and total DNA sequence (>3 billion base pairs (bp) for humans). However, a comparison between the number of genes present in the genomes of different organisms reveals a lack of significant disparity. Estimates are that only 2% of the genomes of complex eukaryotes code for proteins, whereas a much greater amount of sequence is dedicated to regulation of gene expression, accounting for the increased size of the genome and lack of significant increase in gene number (IUM 2012, Int'l Hum Genome Sequencing Consortium 2004). Understanding the sequences that regulate gene expression and how they function in the context of the larger eukaryotic genome remains a major priority for genetics research.

Species	Size of genome (Mb)	Approximate number of genes	References
Eukaryotes			
<i>Arabidopsis thaliana</i> (plant)	125	25,500	AGI (2000)
<i>Caenorhabditis elegans</i> (nematode worm)	97	19,000	CESC (1998)
<i>Drosophila melanogaster</i> (fruit fly)	180	13,600	Adams <i>et al.</i> (2000)
<i>Homo sapiens</i> (human)	3,200	30,000–40,000	IHGSC (2001); Venter <i>et al.</i> (2001)

Species	Size of genome (Mb)	Approximate number of genes	References
<i>Saccharomyces cerevisiae</i> (yeast)	12.1	5,800	Goffeau <i>et al.</i> (1996)
Bacteria			
<i>Escherichia coli</i> K12	4.64	4,400	Blattner <i>et al.</i> (1997)
<i>Mycobacterium tuberculosis</i> H37Rv	4.41	4,000	Cole <i>et al.</i> (1998)

Figure 1. Comparison of the genomes for diverse organisms. (Adapted from Brown, TA, 2002)

Two significant players in the regulation of gene expression include the interaction between “enhancer” and “promoter” sequences. The promoter is a sequence of DNA directly upstream of the gene of interest, where proteins required for the assembly of the RNA polymerase localize to initiate transcription. The RNA polymerase is an enzyme that synthesizes primary RNA transcripts from a template DNA strand. In eukaryotes, transcription occurs in the nucleus and for protein coding genes the primary transcripts are processed into messenger RNA sequences that will be translated into proteins in the cytoplasm by the ribosomes.

Transcription factors are an important class of proteins that bind to DNA in a sequence-specific manner and regulate the occurrence of transcription. These include the general transcription factors which bind to sequences within the promoter. In the absence of these factors binding, RNA polymerase cannot recognize the transcriptional start site of a particular gene, resulting in a lack of gene transcription. One example of a promoter sequence that helps to initiate transcription in eukaryotic genomes is the TATA sequence: 5'-TATAAA-3'. This TATA-like sequence is present in ~24% of gene promoters of

eukaryotes, and contributes to the initiation of transcription by binding the TATA-binding protein (TBP) (Yang, Bolotin, Jiang, Sladek, Martinez, 2007). The TBP starts a cascade of general transcription factor binding events to the promoter region, resulting in ultimate recruitment of the RNA polymerase and the stimulation of RNA polymerase to initiate transcription.

To add further complexity to the process of eukaryotic transcription, additional sequences known as enhancers are required to initiate transcription in specific cell types (for multicellular organisms), under cell environmental conditions, and at certain times during an organisms life. Enhancer regions can regulate the initiation of transcription by binding with various transcription factors that can then interact with the transcription factors that are bound to the promoter. In multicellular eukaryotes, the lack of proportional increase in the number of genes present in the genomes compared to the increase in organism complexity is compensated by the recurrent expression of the same genes in different cell types, life time points, and/or under different environmental conditions. This reuse of genes is facilitated by an increase in the enhancers that can regulate each gene, resulting in either gene activation or gene repression depending on the transcription factors that are bound to the enhancer (Figure 2). Because enhancers contain a multitude of transcription factor binding sites, differential binding of these sites results in distinctive patterns of expression. This varied expression shows the importance of such interactions in that gene regulation dictates the phenotype of the individual.

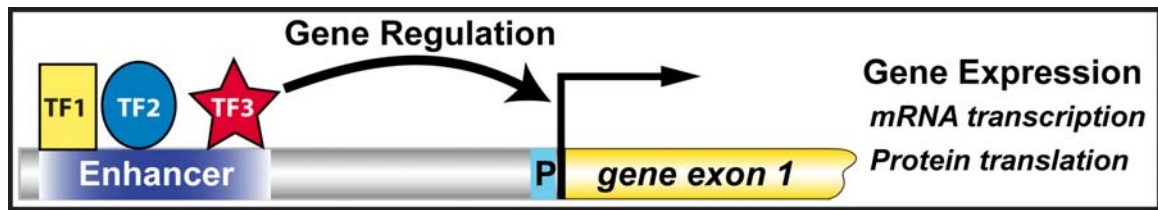


Figure 2. Conceptual overview of eukaryotic gene regulation. Gene expression requires the initiation of transcription downstream of gene promoter regions. For protein coding genes, transcription results in the production of proteins by the translation of mRNA molecules. Transcription is regulated to by that activity of multiple transcription factor proteins (eg. TF1-TF3) that bind to sequence known as enhancers. These transcription factors relay a regulator state to the promoter region of a gene.

While promoters are located just upstream of a gene’s transcriptional start site (Figure 2), enhancers reside in more diverse locations including adjacent or proximal to a promoter of regulation, introns and regions both upstream and downstream of the regulated gene. Moreover, enhancers can be located at great distances from their target promoter, often in closer proximity to non-target gene promoters. In order for these “distal” enhancers to communicate with the promoter, the enhancer must come into close proximity to the promoter, which is thought to occur through a DNA looping event. In this looping model the transcription factor(s) bound to the enhancer interacts directly with a transcription factor(s) bound to the promoter, bringing the two sequences together allows for the activation of transcription (Figure 3). One of the most characterized examples of these long-distance interactions is the regulation of the *Sonic hedgehog* (*Shh*) gene (Lettice et al. 2003). In this particular instance, limb bud expression of the mouse

Sonic hedgehog gene is under the control of an enhancer located over a million base pairs upstream of the promoter (Figure 4), and embedded in the intron of the *Lmbr1* gene that is not even expressed in the limb bud. Thus, enhancers and distally-located promoters must be able to communicate over distances via DNA looping, and these interactions must be encoded within their respective DNA sequences.

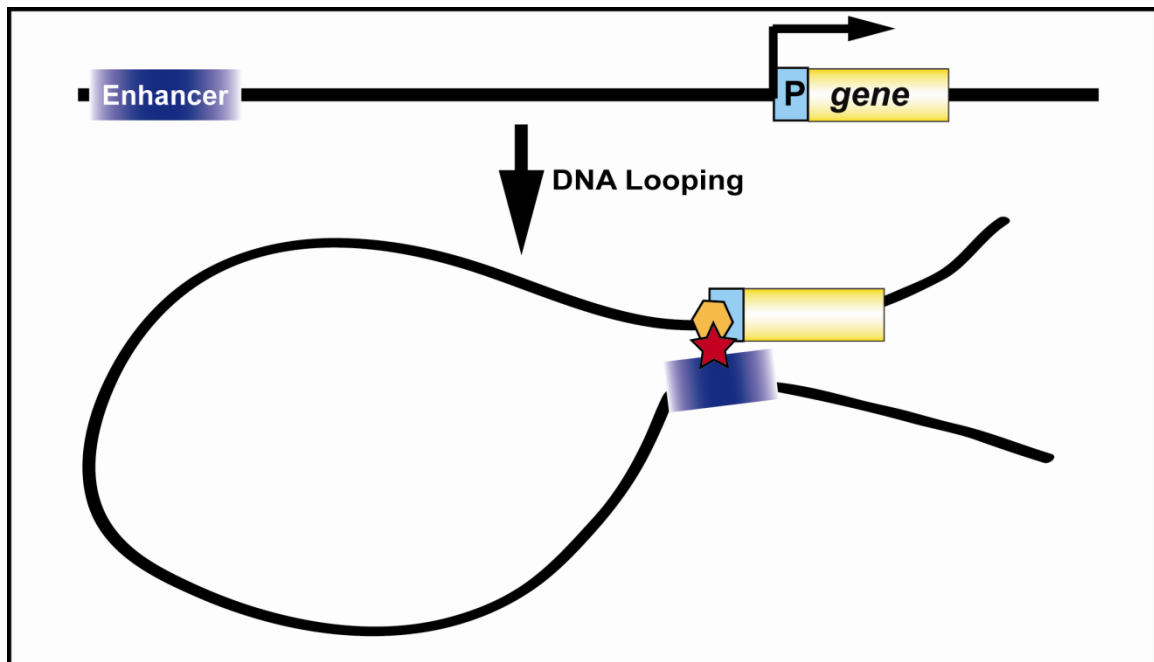


Figure 3. Model for gene regulation by looping interaction between a distal enhancer and a target gene promoter. Star and hexagon shapes represent transcription factor proteins bound respectively to binding sites with the enhancer and promoter.

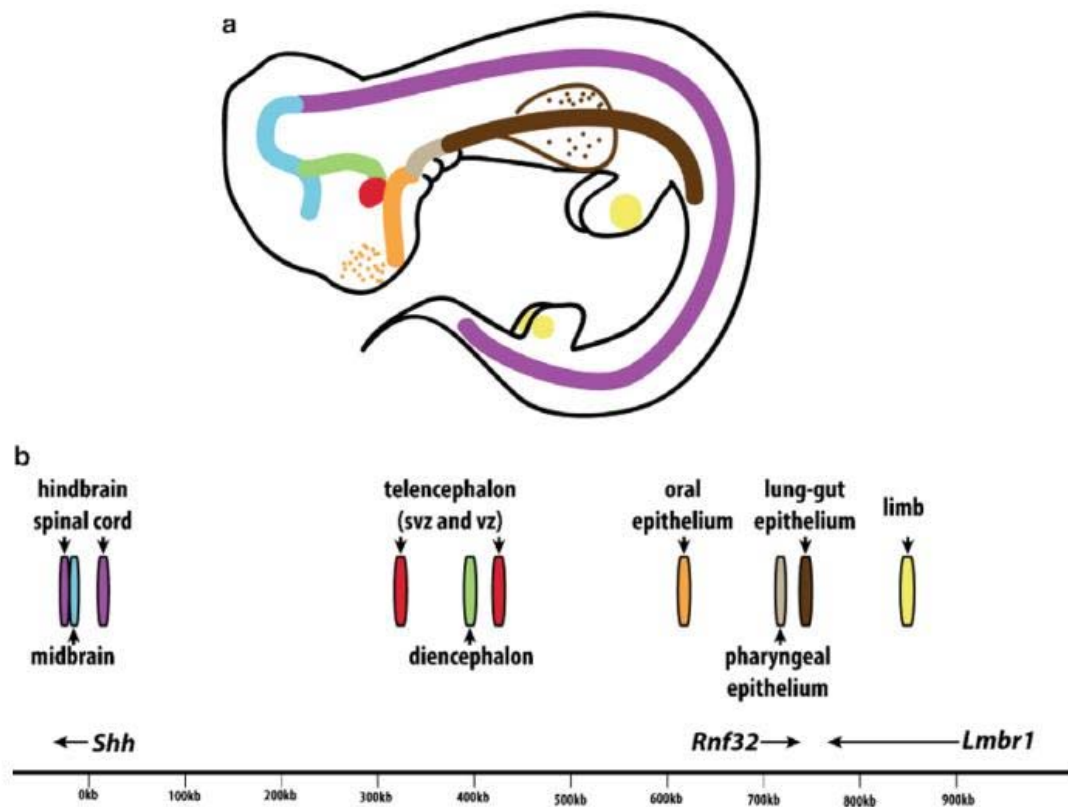


Figure 4. Numerous widely distributed enhancers regulate *Shh* gene expression. (a) Expression of *Shh* during mouse embryonic development. (b) Chromosomal location of enhancers that regulate *Shh* expression. Enhancer color matches the expression pattern represented above. The limb bud enhancer is located in an intron of the *Lmbr1* gene. This Figure was adapted from Williams and Rebeiz (2011).

Recent advances in comparative genomics have allowed scientists to identify conserved non-coding sequences in genomes of related species that may act as “enhancers” to regulate gene expression (Visel, Bristow, Pennacchio, 2007). However, functional studies are needed to determine if the conserved sequences does indeed have

enhancer activity. To determine the function of a sequence of DNA, like the conserved regions in metazoan genomes, a reporter gene can be utilized. The reporter gene is often a gene for a fluorescent protein, usually *Enhanced Green Fluorescent Protein* (EGFP), which reveals the activity of an enhancer *in vivo* (Figure 5). When enhancer sequences are identified, the functional sequences contained within can be identified by introducing mutations to the sequence and seeing whether the mutated sequences results in altered reporter gene expression (Figure 5). Typically, reporter transgenes include one reporter gene adjacent to a heterologous reporter, such as that for the fruit fly *hsp70* gene (Sambrook, Fritsch, Maniatis, 1989). Enhancers are then situated immediately adjacent, or proximal, to the promoter. This reporter transgene conformation eliminates any need for sequence functions that would be involved in spatial regulation through DNA looping.

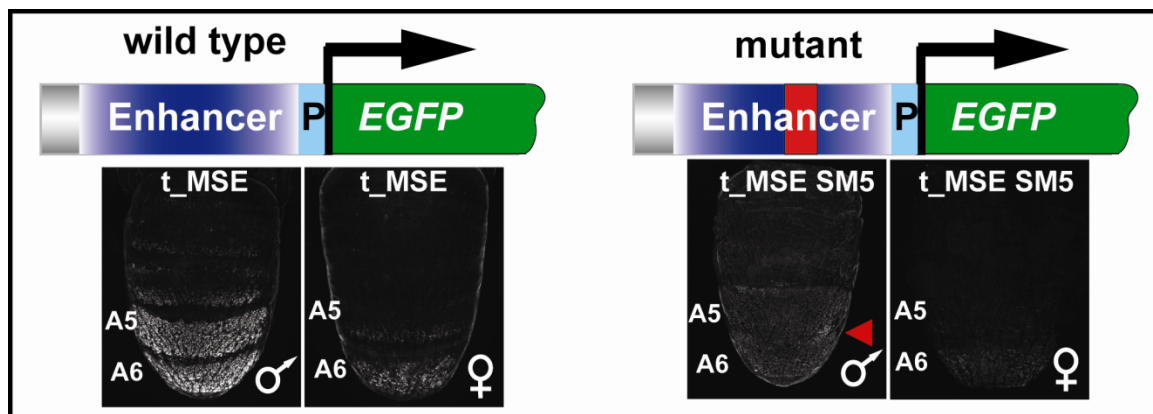


Figure 5. The traditional utilization of reporter transgenes to study gene regulation by enhancers. Enhancers are usually placed immediately adjacent to a minimal sufficient promoter sequence. Enhancer activity can be visualized by monitoring the location, time point, and amount of reporter protein production. The t_MSE enhancer activates EGFP

expression in the pupal abdomen of male *Drosophila melanogaster*. When certain mutations are introduced (t_MSE SM5) reporter expression can be significantly altered indicating the mutated sequence was necessary for the enhancer's gene regulatory capability. Unpublished data kindly provided by Eric Camino.

A recent study has elucidated the specific sequence in the enhancer region that dictates its ability to locate and interact with a gene promoter when the enhancer was placed at a modest distance (~800 base pairs). This enhancer sequence was dubbed a “remote control element” (Swanson, Evans, Borolo, 2010). This study characterized the distance-dependent function of the *sparkling* enhancer on the regulation of the *Drosophila melanogaster Pax2* gene as it functions as a transcription factor that facilitates DNA looping. However, in order to study this distance dependent interaction, two copies of each transgene were created, one proximally located enhancer (~100 base pairs) and one distally located enhancer. While the study provided evidence of gene regulation that was dependent on the distance between the enhancer and the promoter, the multitude of transgenes that had to be created was not ideal. Therefore, my project proposed to develop a transgene system where the proximal and distal relationships occur in the same transgene to simultaneously study the sequences that allow an enhancer to identify and interact with its target promoter. Moreover, this system could conceivably be used to identify the specific remote control sequences responsible for enhancer activity over a distance.

In this transgene system, I will evaluate components of the *Drosophila melanogaster* *bric-à-brac* (*bab*) locus, which has the tandem duplicate *bab1* and *bab2* genes that are expressed in many cell types and regulated by many enhancers (Couderc, Godt, Zollman 2002; Williams et al. 2008). The *bab* gene locus is a large, spanning ~160,000 base pairs, with numerous enhancers that are located throughout the gene locus (Figure 6). I will specifically study the enhancer referred to as the dimorphic element (DE), as it directs *bab1* and *bab2* expression in the posterior female abdomen respectively located ~13,000 and ~90,000 bp from the *bab1* and *bab2* promoters (Figure 6A, enhancer “6”; and 6G, reporter transgene activity). This enhancer is bound by two known transcription factors: Abdominal B (AbdB) and Doublesex (Dsx). However, the dimorphic element enhancer contains extensive sequence that has been evolutionarily conserved, suggesting that there are additional transcription factors that bind the dimorphic element whose identities remain unknown (Rogers, Salomone, Tacy, Camino, Davis, Rebeiz, Williams, 2013). These additional transcription factors may be involved binding to a remote control element that functions in a looping interaction with the *bab* gene promoters.

My thesis aims to reveal the spatial component of enhancer-promoter interaction by using a dual reporter gene vector (pRLGL2 or also known as “Red Light, Green Light”) to differentiate between proximal and distal enhancer-promoter communications. Using the *bab* locus in *Drosophila melanogaster* as my model system, I first created constructs of the vector to see whether a heterologous promoter (*hsp70*) would lose its ability to communicate with the dimorphic element - an endogenous *bab* enhancer. Second, I sought to test whether putative *bab1* and *bab2* promoters would communicate

with the distally located dimorphic element. Since the endogenous promoters communicate with the dimorphic element *in vivo*, this replacement was expected to result in a recapitulation of long-distant communication between this enhancer and the endogenous promoter. This novel reporter transgene system seems to have the potential to reveal and study the spatial component that is often overlooked in gene regulation. This system should be useful for investigating these distal interactions in other genes as well, resulting in a deeper understanding of gene regulation in general. Because enhancers and promoters are present in all animals, including humans, this system can be applied to study similar long distance interactions involved in gene regulatory complexes that give rise to pathophysiological conditions.

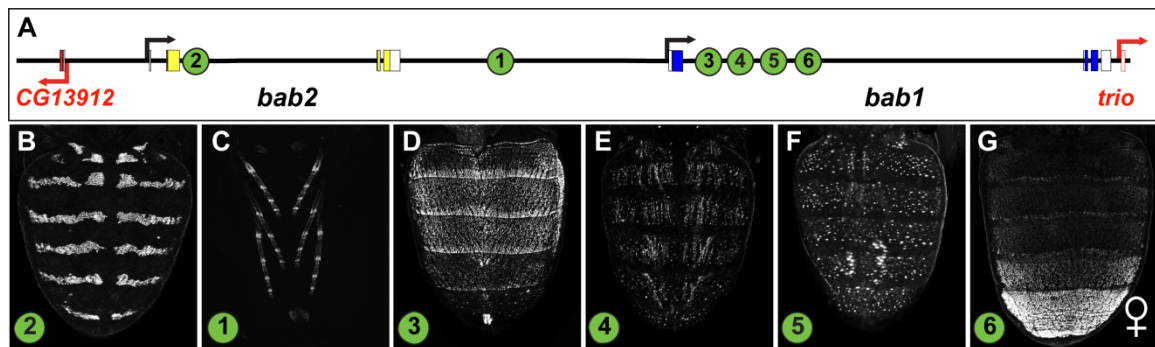


Figure 6. Gene regulation at the *bab* locus. (A) Enhancers, represented as numbered circles, are distributed throughout the *bab* locus. Promoter positions for the *bab1* and *bab2* genes are represented as black arrows, and red arrows for the flanking *CG13912* and *trio* genes. EGFP expression patterns driven by the enhancers for the (B) oenocytes, (C) leg tarsi, (D) anterior abdomen, (E) abdominal muscles, (F) abdominal sensory organs, and (G) female posterior abdomen.

CHAPTER II: MATERIALS AND METHODS

A. Vector Design

The Red Light Green Light (pRLGL) vector was created as a dual reporter gene vector allowing differentiation between proximal and distal enhancer-promoter communications (Figure 7). The original vector was designed using an mS3aG vector (John Butts, unpublished) containing the *hsp70* gene promoter (referred to as “*hsp70p*”) and the coding sequence for the *Enhanced Green Fluorescent Protein* (*EGFP*). A first generation vector called “pRLGL was made by synthesizing a EcoRI-AscI fragment (Appendix M) which contains an *hsp70p-DsRed2* gene cassette, and a AscI-AgeI fragment that contains the dimorphic element, a 1 kb spacer sequence and a *hsp70p* (Appendix M). An EcoRI and AgeI fragment was removed from mS3aG inserting and the synthesized pieces were added in its place. This placed the second *hsp70p* next to the *EGFP* gene of mS3aG. *DsRed2* (technically DsRed.T4.NLS; (Barolo, Castro, & Posakony, 2004)) codes for a nucleus-localized Red Fluorescent Protein (RFP). In this vector, *EGFP* was distal to the dimorphic element. We found that the proximal *DsRed2* gene made substantial amounts of variant green fluorescent protein form (Eric Camino, unpublished results). This limited our ability to distinguish signal from the *EGFP* gene and thereby from the distally-located promoter.

From the original pRLGL vector, we decided to make a second generation vector where *DsRed2* and its associated promoter would be located distal to the dimorphic element (Figure 7). We call this vector pRGL2, and this vector was the substrate for my thesis work. To make this vector, a StuI-AgeI fragment was removed from pRLGL and

replaced with a custom synthesized fragment (Appendix O). This fragment contained a 2.0 kb spacer sequence position to make the dimorphic element distal to the *hsp70p*-*DsRed2* gene. The final 14,800 base pair vector contains the *EGFP* gene that will be expressed in the pattern directed by the proximal dimorphic element, and the *DsRed2* gene that will be expressed from a promoter distal to the dimorphic element. The distal promoter is referred to as the “test” promoter as different promoter sequences will be evaluated in this position.

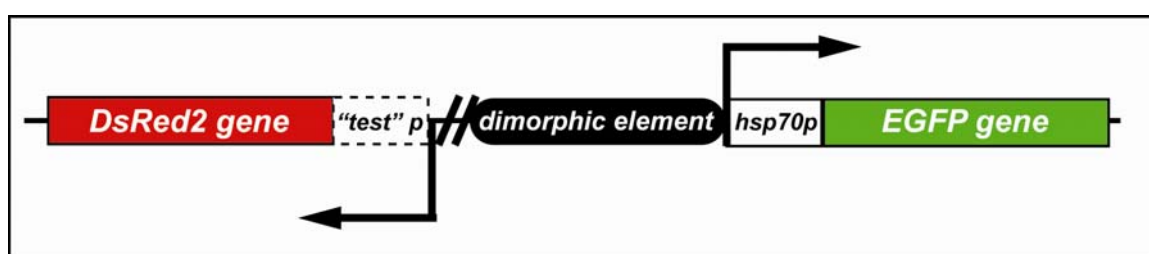


Figure 7. Spatial organization of pRLGL2 vector. Two fluorescent protein gene vector that includes two promoters and a single enhancer (in this study the dimorphic element). A spacer sequence is placed between the slashes to make the promoter regulating the *DsRed2* gene expression distal from the dimorphic element. Unique restriction enzyme sites flank the “test” promoter sequence and the dimorphic element so that variant promoters and enhancer sequences can be substituted using endonuclease enzymes that have cut sites in the vector.

B. Vector Preparation- altering the length of the spacer sequence

The original goal was to make spacers of variable lengths to test for effects on expression from the distal promoter. This included spacers of 0 kb, 0.25 kb, 0.5 kb, and

1kb in addition to the vector with the 2.0 kb spacer described above (Figure 8). Due to time and cloning scheme complications, we limited this study to 0.5 kb and 1.0 kb.

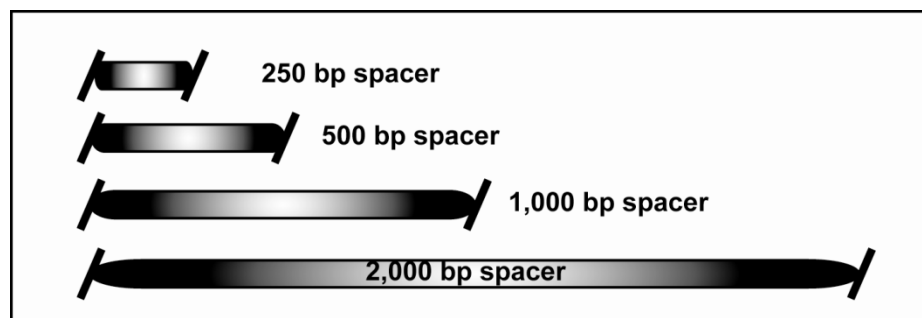


Figure 8. Scheme for varying spacer size between dimorphic element and the distal promoter regulating DsRed2 expression.

B.1: 0KB Spacer

To make the 0 kb spacer construct of the pRLGL2 vector, the pRLGL2 vector with the 2 kb spacer was propagated in bacteria following transformation in competent *E. coli* cells (Appendix E). Cultures from transformed colonies were incubated with LB medium-Ampicillin (Appendix D) at 37°C for 18 hours in snap-cap tubes with rotation. These samples were purified using a QIAgen Mini Prep procedure (Appendix F), resulting in samples of isolated pRLGL2-2kb vector. The samples were then digested sequentially using the restriction enzymes BamHI-HF and Bgl II that flank the 2.0 kb spacer (Figure 7 and Figure 8).

First, 4ul of BamHI-HF was used to digest 55ul of pRLGL2 using 10ul of NEB Buffer 4, 10ul 10XBSA, and 21ul of Milli-Q to buffer. This reaction was digested at

37°C for 30 minutes. Then, the enzymes and buffers were removed from the reaction using an ethanol precipitation procedure (Appendix G) in preparation for the Bgl II digest. Finally, 2ul of Bgl II was used to digest the 50ul sample of purified BamHI-HF digested DNA using 10ul of NEB Buffer 3, 10ul 10XBSA, and 8ul of Milli-Q to buffer. This reaction was digested at 37°C for 22 hours. The fully digested pRLGL2 vector was run in a 0.7% gel (Appendix C), and the vector lacking the 2.0 kb insert was purified using the QIAgen large band gel purification procedure (Appendix J). This vector backbone lacking the spacer was then ligated together using T4 DNA ligase and standard protocol. The ligation products were transformed into competent *E. coli* cells. A plasmid vector lacking the spacer (pRLGL2-0kb) was confirmed by DNA sequencing (Eric Camino, personal communication).

B.2: 0.5KB and 1KB Spacer vectors

Stage 1: Amplification and Purification of 0.5kb insert

The 0.5 kb and 1.0 kb spacers were amplified from the 2.0 kb spacer-containing vector with a Phusion DNA polymerase PCR protocol (Appendix A), using the primers indicated in Table 1. To remove the enzymes and buffers from the PCR amplified DNA, a PCR purification protocol was performed (Appendix H). Following the purification, the DNA was digested with the restriction endonucleases BamHI-HF and AscI to create the 5' and 3' overhangs for ligating into the pRLGL2 vector that had the spacer removed by AscI and BamHI digestion (discussed below). The enzymes and buffers from the

digested 0.5 kb and 1 kb inserts were removed from the samples using gel electrophoresis followed by a small DNA fragment purification protocol (Appendix I).

Table 1. Primers for amplifying smaller spacer sequences.

Primer	Sequence	Restriction Site
RLGL2 0.5K spc Fwd1	TTTCCGggatccCGTGTGGGATACTGGTACTGG	BamHI
RLGL2 .5K spc Rev2	TTGCCggcgcgccCGGAGTTGGCTGTAGCTATCTTGC	AscI
RLGL2 1K spc Fwd1	TTTCCGggatccCGTGTGGGATACTGGTACTGG	BamHI
RLGL2 1K spc Rev2	TTGCCggcgcgccGTGTATCCGTCCCAGTACCTCG	AscI

Stage 2: pRLGL2-2kb Vector Digestion

The pRLGL2 vector for the 0.5 kb and 1.0kb inserts was prepared from the original 2 kb spacer construct. 50ul of the pRLGL2-2 kb vector was digested with 1ul of each restriction endonuclease- BamHI-HF and AscI- using 5ul of NEB Buffer 4, 5ul of 10XBSA, and 18ul of Milli-Q as a buffering solution. This digested sample was then run in a 1% gel for 30 minutes using a large comb and purified using a large DNA fragment purification (Appendix J), resulting in a construct of the RLGL vector with overhangs cut in accordance with the restriction enzyme cleavage sites.

Stage 3: Ligation and Bacterial Transformation

With the 5' and 3' overhangs on both the insert and the vector cut with the same restriction endonucleases (AscI/BamHI-HF), the 0.5 kb and 1 kb spacers could be ligated into the pRLGL2 vector using a T4-ligase-mediated ligation (Appendix D). This ligation was then transformed using competent *E. coli* cells using a standard bacterial transformation protocol (Appendix D).

Stage 4: Mini Prep and Genomic Sequencing

Colonies from the bacterial transformation were incubated with LB medium-Ampicillin at 37°C for 18 hours in snap-cap tubes with rotation, and the vectors containing the 0.5kb and 1kb inserts were isolated using standard mini prep procedures (Appendix F).

Following the plasmid mini preps, the DNA concentrations were quantified using UV Spectrophotometric analysis with a nanophotometer (Denville Scientific), and sent to DNA analysis LLC for sequencing to ensure the correct DNA pieces had been cloned into the vector (Sequencing confirmed by Eric Camino).

B.4: 2 kb Spacer vector

To check the 2kb spacer construct for the pRLGL2 vector, the Gen Script© stock was first amplified using bacterial transformation spot plating (Appendix E), then incubated and purified using the same procedures from the 0kb spacer isolation. 5ul of the sample was then digested with 1 ul NEB Buffer 4, 1 ul 10XBSA, 2.5 ul Milli-Q, and 0.25 ul of each of the restriction enzymes- StuI and AgeI- in accordance with the restriction cleavage sites noted in the RLGL vector design (Appendix O). This test digest was then

loaded into a 0.7% agar gel, and run for 30 minutes to show that vector maintained integrity throughout the amplification. Finally, the vector concentration was quantified using UV Spectrophotometric analysis with a nanophotometer (Denville Scientific), and sent to DNA Analysis LLC. to be sequenced (Confirmed by Eric Camino).

C: Vector Preparation- Promoters

Once the pRLGL2 vector constructs with different lengths for spacer sequences was established, the original promoter from the 2kb construct (*hsp70*) was removed and replaced by various promoters from the endogenous *bab* locus. This part of the experiment addressed the whether promoter sequences effect communication with a distal enhancer (dimorphic element). This test included the putative *bab1* and *bab2* promoters (*babp1* and *babp2*) using the 2kb spacer construct. One other construct was created to act as a control, in which the *bab2p* and the *hsp70* were both present in *cis* to each other. In this part of the experiment, we expect the endogenous promoters of recapitulate expression of the RFP in a type 1 pattern (see results chapter). This would be evidence that a remote control element or elements in the dimorphic element had selectivity in its promoter interactions.

Stage 1: PCR Amplification of Promoters

Before the vector preparation with variable promoters, it was necessary to isolate these promoters (*bab1p* and *bab2p*) from the endogenous *bab* gene locus, and amplify the

quantity for further experimentation. These promoters were amplified using Phusion PCR protocol (Appendix A) using promoter-specific primers as indicated in Table 2.

Table 2. Primers to PCR bab gene promoters for cloning

Primer	Sequence	Restriction Site
b1p Fwd RLGL	GTGTCCAGTGTTATTCAGTGTGTGC	StuI
b1p Rvs RLGL	CATCAACGTTGCCCCTCAGCTTTCC	BamHI
b2p Fwd RLGL	GTGGCCGTAAAAACAAACG	StuI
b2p Rvs RLGL	GTCGGCAGCGCGGCGTCGAG	BamHI
b2p+HS70 Fwd RLGL	GTGGCCGTAAAAACAAACG	StuI
b2p+HS70 Rvs RLGL	GTCGGCAGCGCGGCGTCGAG	BamHI

The Phusion PCR pieces were then run on a 0.7% gel for 30 minutes to show that the correct sequences were amplified (Figure 9). According to the primers used to isolate the promoters, the *bab1p* should have been about 300 bp in length, and the *bab2p* should have been about 200 bp in length, which can be visualized in the image below. PCR fragments are separated in a gel based on their size. The base pair ladder (2 log ladder, New England Biolabs) that is shown next to the gel image is a generic indicator that is the basis of which the PCR fragments in the gel are sized.

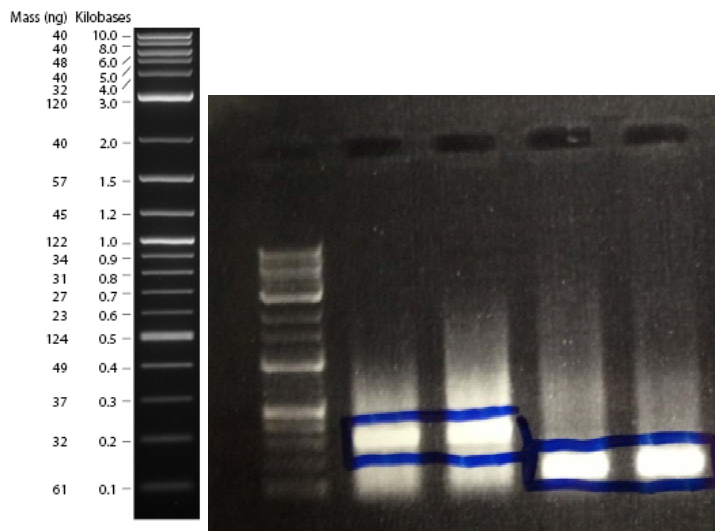


Figure 9. PCR-amplified promoters for the *bab1* and *bab2* genes.

Stage 2: Promoter Digestion

Once the promoters were amplified and isolated using PCR and gel electrophoresis, the fragments were purified using a small DNA fragment purification (Appendix I) and prepared for insertion into the 2kb spacer pRLGL2 vector. All three promoters were digested with restriction endonucleases BamHI-HF and StuI to create the 5' and 3' overhangs for insertion into the pRLGL2 vector. For the *babp1* and *babp2*, the insert contained the 200-300bp promoter sequence with the BamHI-HF and StuI overhangs. The *babp2+hsp70p* insert contained an ~500bp sequence that included both the *bab2* promoter and a heat shock promoter. Because the *bab2p* and the *hsp70p* are combined into one insert with the BamHI-HF and StuI overhangs, the same pRLGL2 vector was used for insertion of all three promoter constructs.

Stage 3: RLGL Digestion

The 2kb construct of pRLGL2 was used as the template for the variable promoter constructs. The same enzymes- BamHI-HF and StuI- that were used to prepare the promoter inserts were used to open the vector to create complimentary overhang sequences to the promoter inserts. The restriction digest resulted in a vector in which the *hsp70* promoter that was located proximal to the DsRed2 was removed using BamHI-HF and StuI, leaving overhangs that are complimentary to the promoter insertions. Because the promoter inserts were so small, an additional step was taken to ensure the insertions are accurately introduced to the vector: SAP. SAP is an enzyme that catalyzes the removal of 5' phosphate from the DNA of interest; since phosphatase-treated fragments lack the 5' phosphoryl termini required by ligases, they cannot self-ligate. This step increases the probability that the vector will ligate to the promoter insert during ligation, and avoid ligating to its own complimentary overhang.

Stage 4: Ligation and Bacterial Transformation

The isolated promoter sequences were then ligated, using T4 ligase, into the pRLGL2 vector that had been prepared using restriction endonucleases BamHI-HF and StuI, in accordance with the with the restriction cleavage sites noted in the pRLGL2 vector design (Appendix O). The SAP step in the pRLGL2 vector preparation protocol facilitated this ligation by ensuring that the vector does not ligate with itself, so the promoter piece is more likely to be incorporated. The ligated vector and promoter insert was then transformed using z-competent *E. coli* cells from the Williams Lab stock using standard bacterial transformation protocol (Appendix D).

Stage 5: Mini Prep and Genomic Sequencing

Colonies from the bacterial transformation were incubated with LB medium-Ampicillin at 37°C for 18 hours in snap-cap tubes with rotation, and the vectors containing the 0.5kb and 1kb inserts were isolated using standard mini prep procedures (Appendix F).

Following the mini prep, the promoter inserts were checked for correct insertion using a Go-Taq polymerase (Promega Inc.) PCR (Appendix B). The primers used for this sequence verification were chosen so that if the sequence of the promoter is present, the PCR will be successful, but if the sequence was not present, the PCR amplification would not occur. For this purpose, DsRed2 Fwd was used as the forward primer, and *bab2p* Rvs was used as the reverse primer for the *bab2p* and the *bab2p+hsp70* construct as indicated in Table 2. For *babp1*, the forward primer was the DsRed Fwd, similar to the *bab2p/bab2p+hsp70p* constructs, and the reverse primer was *the bab1p* Rvs (Table 2).

D: Confocal Microscopy

D.1: Transgenics and Homozygous Fly Preparation

The transgenic flies (with genomically integrated pRLGL2 vectors) that were created for this experiment were generated by Best Gene Inc, using the vectors that were prepared in part A-C. Transgenes were integrated into the attP40 genomic landing site (Markstein, Pitsouli, Villalta, Celniker, & Perrimon, 2008) using phiC31 integrase methods for site specific integration (Groth, Fish, Nüsse, & Calos, 2004; Williams et al., 2008). Once the transgenic flies were obtained, the lines needed to be made homozygous for the transgenes that was achieved by selecting and mating flies with dark red eye

colors (Rogers & Williams, 2011). The transgenic *Drosophila melanogaster* flies were grown at room temperature (21°C) in vials with a pre-made mixture of sugar food (Appendix L) (Salomone, Rogers, Rebeiz, & Williams, 2013). Specimens were grown for 12-14 days before pupae were selected for dissection. The pupal stage for analysis was equivalent to ~80 hours after puparium formation for *Drosophila melanogaster* pupae cultured at 25°C.

D.3: Dissection

First, packing tape was placed onto a plastic dissection board and fastened in position, sticky-side up using additional pieces of Lab Label tape at the top and bottom ends. The flies that appeared to be in the ideal stage of pupae development were selected and placed on the sticky-side of the tape. Using a pair of forceps, the pupa covering was removed from the specimens, and the remaining fly body was removed from the tape, and transferred to a microscope slide to be viewed under the confocal microscope.

D.4: Laser Scanning Confocal Microscopy

EGFP and DsRed2 expression were visualized in the dorsal abdominal epidermis using the University of Dayton Biology department's confocal microscope (Olympus FV1000) (Rogers, Williams, 2011). With a confocal microscope (Fluoview 1000) with the 10X objective, a series of images were taken of each transgenic pupae along the z-axis at increments of 10 microns. These series of images were taken at the settings listed in Table 3 and Table 4, and at a Kalman line averaging setting of 3.

Table 3. Confocal microscope settings used for the imaging of EGFP expression for Red Light Green Light transgenic pupae.

Laser	Laser Output	HV	GAIN	OFFSET	Emission Wavelengths Collected
700	15%	700	1	1	500-535

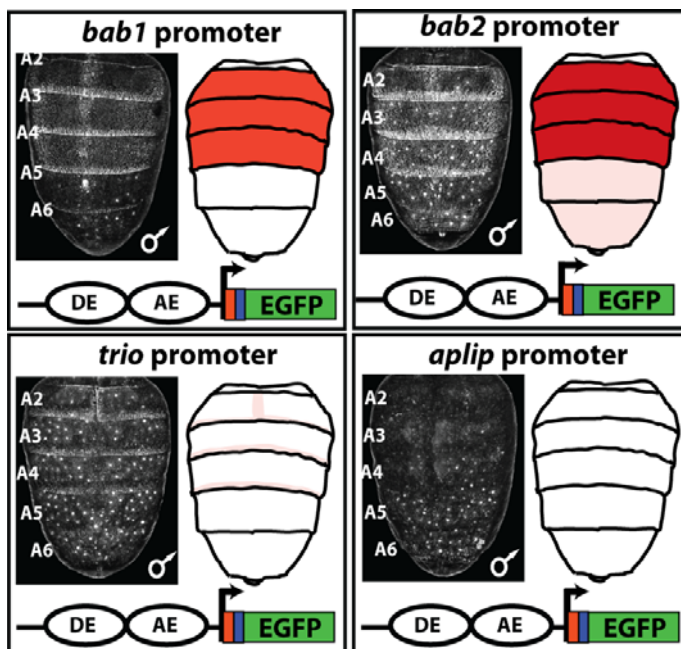
Table 4. Confocal microscope settings used for the imaging of DsRed2 expression for Red Light Green Light transgenic pupae.

Laser	Laser Output	HV	GAIN	OFFSET	Emission Wavelengths Collected
650	10%	650	1	1	580-625

Using Fluoview software, a projection image was exported in Tagged Image File Format (TIFF) after the series of images were completed. Multiple pupae were analyzed for each pRLGL2 transgene in order to make sure results were reproducible. Images were processed using Adobe Photoshop, and each image received the same series of modifications.

CHAPTER III: RESULTS

My research aimed to develop a transgenic system that will help identify regulatory inputs that are necessary for enhancer-promoter interactions. This project will address several hypotheses. The first is that gene expression output is inversely related to the distance between an enhancer and a heterologous promoter (from a gene not regulated by the enhancer *in vivo*). The second hypothesis is that native promoters will reestablish



communication with a remote enhancer. We demonstrated for the *bab* gene locus that the dimorphic element and anterior element enhancers more robustly activated reporter gene expression when the promoter included sequences for the putative *bab1* and *bab2* gene promoters than when possessing a putative sequence for either the *trio*

and *aplip* gene promoter (Figure 10, left). The last hypothesis that will be addressed is that a screen of mutations for an enhancer will identify a “remote control element” or “RCE” (Swanson, Evans, & Barolo, 2010) as an enhancer subsequence that is required for promoter communication over a distance. To test these hypotheses, this project targets three aims:

Aim 1: Test effects of enhancer-promoter spacing on reporter gene expression.

Aim 2: Compare abilities of endogenous and heterologous promoters to communicate with a remote *bab* locus enhancer.

Aim 3: Screen a set of enhancer mutations to find sequences that function as a promoter remote control element.

Results for Aim 1: Spatial

To determine the effects of proximity on enhancer-promoter communication, alternate transgenes were constructed, where the space between the enhancer and the *DsRed2* gene (encoding a Red Fluorescent Protein or RFP) are 0, 250, 500, 1000, and 2000 bps (Figure 11).

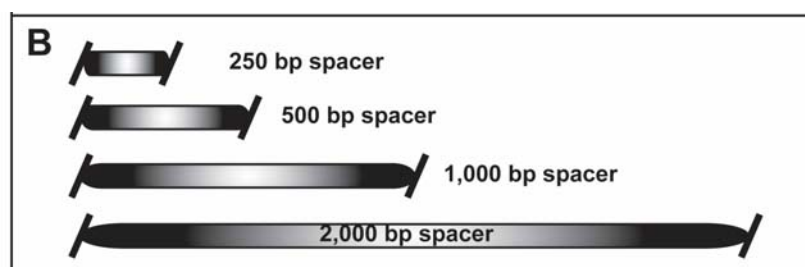


Figure 11: Spacer constructs for insertion into pRLGL2

I anticipated that the spacer sequence manipulation would result in one of four possible outcomes (Figure 12). A Type 1 outcome is where increasing the distance between an enhancer and a promoter has no effect of reporter expression. Here, RFP and EGFP would be expressed similarly, irrespective of spacer length. Alternate outcomes that I favored, were where reporter (RFP) expression is either reduced (Type 2 outcome) or lost altogether (Type 3 outcome) as distance is increased between the enhancer and the distal heterologous promoter. A Type 4 outcome is one where both distal and proximal

reporter expression is lost (Figure 12). My expectation was that the pRLGL2 vector containing the dimorphic element enhancer and *hsp70* promoters at each fluorescent protein genes would result in a Type 1 outcome when spacing included 0 kilobases (kb) in length, a Type 2 outcome when spacing was increased 0.5-1.0 kb in length, and Type 3 outcome when spacing was increased to 2 kb. This set of outcomes would be consistent with my hypothesis that an enhancer, the dimorphic element, cannot communicate well with a distal heterologous promoter (*hsp70*). The lack of communication might be due to an inability for this enhancer/promoter pair to interact with each other.

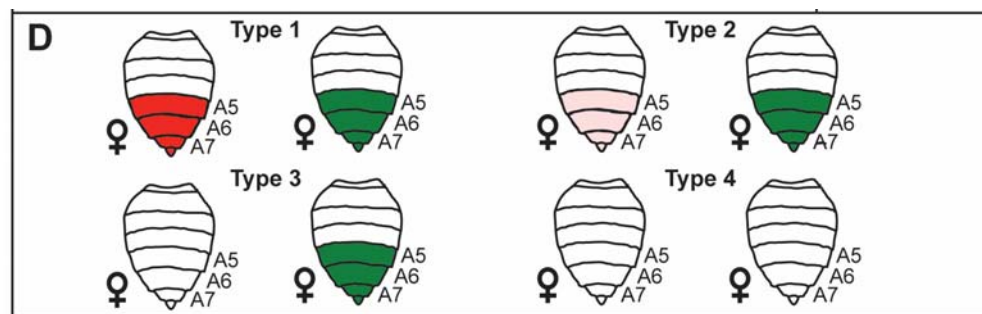


Figure 12: Potential spatial outcomes using RLGL

I. 0kb and 2kb Spacers

The first transgenic flies tested possessed the pRLGL2 transgene with either the 0 kb or 2 kb spacer sequence and the heterologous promoter (*hsp70*) for initiating transcription from the *DsRed2* reporter gene. The preliminary spatial data showed promising results in that RFP and GFP expression was robust in the female A5 and A6 abdominal segments when the 0 kb spacer sequence was present (Figure 13). This was consistent with my hypothesis as it reflected a “Type 1” expression pattern because the

dimorphic element enhancer equidistance to the heterologous promoters upstream of both the *DsRed2* and *EGFP* reporter genes. Much to my surprise, though, robust expression of RFP was seen when the 2 kb spacer was situated between the dimorphic element and the *hsp70* for the *DsRed2* gene (Figure 13). This result shows that the *hsp70* promoter was in fact capable of communicating with the dimorphic element over a distance of 2 kb, despite that this is not the endogenous promoter for the dimorphic element. This RFP quantity appears slightly reduced compared to that when the 0 kb spacing was used. Thus, this outcome can be considered consistent with either a Type 1 or Type 2 outcome. Regardless, it is clear that for this enhancer/promoter combination that communication can occur over a spacing of 2,000 base pairs.

II. 0.5kb and 1kb Spacers

The next transgene construct to be evaluated was that for the 0.5 kb spacer sequence. With these constructs, I expected to see a Type 2 outcome, where GFP remains robustly expressed due to its close proximity between the *EGFP* gene promoter and the dimorphic element, whereas RFP expressed would be attenuated due to the modest 500 base pairs of sequence separating the dimorphic element from the *DsRed2* gene promoter. However, what we saw was more consistent with a Type 1 outcome, which is the outcome we observed for the 2 kb spacer transgene (Figure 13). Due to complications with cloning and the occurrence of Type 1 outcomes for the 0, 0.5, and 2 kb spacer-bearing transgenes, I decided to focus my efforts on Aim 2 and Aim 3. Hence, the data for this construct remains a work in progress at the time of writing my thesis.

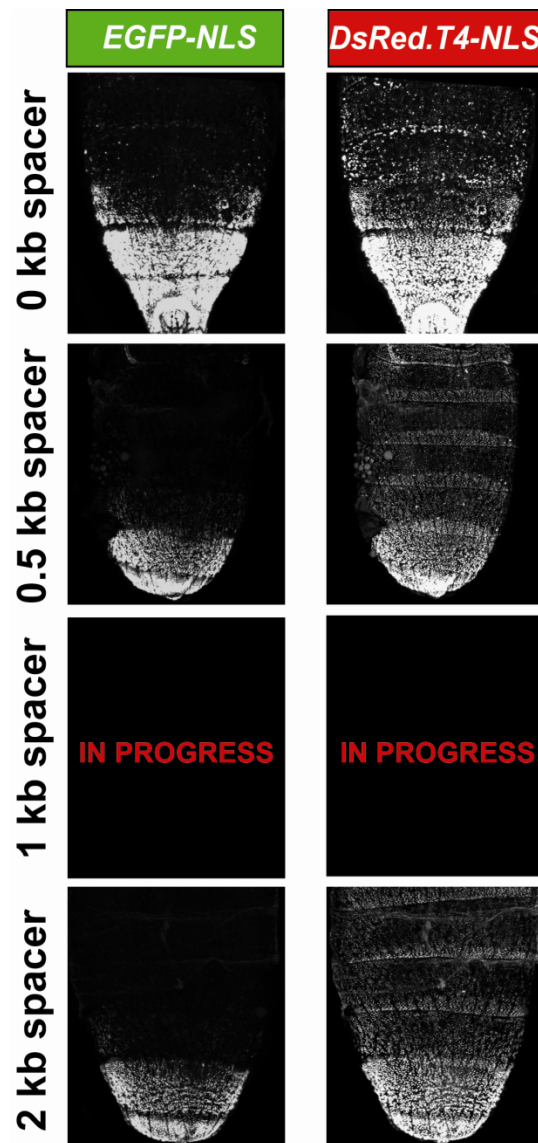


Figure 13. The effects of spacing on interactions between the dimorphic element and the *hsp70* promoter. In the pRLGL2 transgene construct both the *EGFP-NLS* gene (encoding GFP protein) and *DsRed.T4-NLS9* (a.k.a. *DsRed2*) (encoding RFP) are expressed from separate *hsp70* promoters. The effect of spacer sequences inserted between the dimorphic element and the promoter for the *DsRed2* gene was monitored by evaluating the amounts of GFP and RFP proteins produced. The results demonstrate that the dimorphic element can regulate the distal *hsp70* promoter over a distance of 2 kb.

Results for Aim 2: Promoters

According to my original hypothesis, increasing the distance between the *hsp70* promoter controlling transcription initiation of for the *DsRed2* gene would result in a loss of transcriptional activation from the more distal located dimorphic element. In anticipation of this outcome, I decided to test whether the putative promoters for the *bab1* and *bab2* would restore communication with the dimorphic element or embolden this communication. The specific transgene construct that I modified here was the pRLGL2 with the 2 kb spacer.

The *hsp70* promoter was the heterologous promoter that I suspected would not be able to communicate with the dimorphic element enhancer from the *bab* locus. To my surprise, the *hsp70* promoter was able to communicate with the dimorphic element enhancer, despite its heterologous nature, yielding a Type 1 pattern of RFP and GFP expression (Figure 14).

The putative promoter sequences for *bab1* and *bab2* were isolated and substituted in place for the *hsp70* promoter adjacent to the *DsRed2* gene. The *bab1* promoter (*bab1p*) sequence was about 250 base pairs in length, while the *bab2* promoter (*bab2p*) was about 150 base pairs in length. I suspected that these two regions would possess sequences that could physically with the dimorphic element enhancer *in vivo*. Thus, I suspected these promoters would communicate with the dimorphic element over the 2 kb spacer sequence leading to robust RFP production. To my surprise, I found that proximal expression of GFP was unaltered by swapping the distal promoter but the expression of RFP from the distal *DsRed2* gene was markedly reduced (Figure 14). This outcome

suggests that the putative *bab1* and *bab2* promoters fail to communicate with the dimorphic element and/or fail to initiate transcription of the *DsRed2* gene.

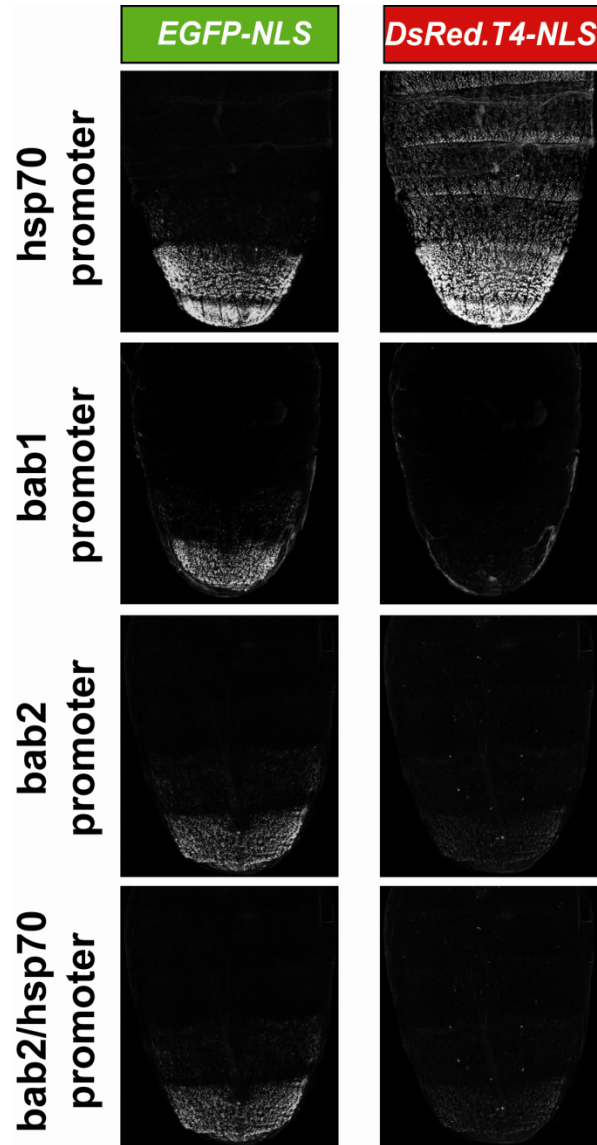


Figure 14. The effects of promoter identity of distant communication between the dimorphic element and a distally-positioned gene promoter. In the pRLGL2 transgene construct both the *EGFP-NLS* gene (encoding GFP protein) and *DsRed.T4-NLS9* (a.k.a. *DsRed2*) (encoding RFP) are expressed from separate promoters. In each case above, GFP expression is controlled by the *hsp70* promoter, whereas RFP expression is

controlled by the promoter listed to the left. The dimorphic element is separated from the distal promoter by a distance of 2 kb.

A priori, it seemed possible that the putative *bab* promoter sequences I isolated might contain sequence needed to interact with the dimorphic element but lack sufficient sequences to initiate transcription. For this reason I created a fusion promoter that contained a 5' *bab2p* followed by a 3' *hsp70* promoter. What I observed was a result intermediate between that seen for the *hsp70* and *bab2p* in isolation. Here, expression of the RFP was moderate (Figure 14). This suggests that the presence of the putative *bab2* promoter sequence was antagonizing the functionality of the *hsp70* promoter that was positioned in *cis*.

The expression outcomes observed when modifying the length of spacer sequence and the promoter identity were surprising and will be discussed in the discussion section.

CHAPTER IV: DISCUSSION

Investigating the mechanisms that drive variations in gene expression requires an understanding of the temporal and spatial aspects through which genes are regulated. Exploring these regulatory mechanisms is reliant on the current understanding of the gene landscape, including the gene promoter and the enhancer regions. Because enhancers can be located in variable distances from the gene promoter(s) for which they regulate transcription, a model system is needed to study so called long-distance interactions between enhancers and promoters. Such a system would allow scientists to develop a deeper understanding of gene regulation in a 3-dimensional nucleus. My thesis research aimed to develop such a model to study these distal interactions by employing the use of a transgene vector with two fluorescent proteins, each associated with a promoter sequence which interacts with a common enhancer. In this case, we used the *bab* gene as our model gene locus, and one of its enhancers: the dimorphic element (Williams et al., 2008). While this system was successfully created, the experimental outcomes suggest several of our initial expectations were incorrect.

Conclusions of the effects of spacing between an enhancer and promoter

I hypothesized that when a heterologous promoter was present in the vector, as the space between the endogenous enhancer and the heterologous promoter increased, there would be a decrease in the interaction between the promoter and the distal enhancer. The results from the spacer sequence data in the pRLGL2 vector (Figure 13)

showed an unexpected trend, in that the heterologous promoter, *hsp70*, was able to communicate with the dimorphic element enhancer over the distance of 2 kb even though the *hsp70* promoter was not an endogenous promoter. Several scenarios were considered to explain this phenomenon, including the efficacy of the *hsp70* promoter as a heterologous promoter and the spacer distance required to disrupt the enhancer-promoter interaction.

Although previous research suggested that this *bab* enhancer showed selectivity for endogenous gene promoters (Figure 10), in the context of Red Light Green Light, *hsp70* was an effective promoter both proximally and distally. This result may have been due to the efficiency of *hsp70* as a promoter; because this promoter is used as a reliable source to drive fluorescent protein expression in most vector constructs (Rebeiz, Castro, Liu, Yue, & Posakony, 2012; Rebeiz & Williams, 2011; Shirangi, Dufour, Williams, & Carroll, 2009; Swanson et al., 2010; Williams et al., 2008), it may just be too efficient of a promoter to be affected by a 2 kb spacer sequence. Succinctly put, the promoter I chose is an outlier and not representative of promoters in general. Perhaps future experiments can incorporate a different heterologous promoter, which would be more likely to be affected by a 2 kb spacer sequence. Two possible candidates are those found in the *trio* and *aplip* genes, which I evaluated in Figure 10 and that flank the *bab* locus.

A second surprise in my results was that the 2 kb spacer was not a sufficient length to suppress expression from the distal *hsp70* promoter that controls transcription of the *DsRed2* gene. *In vivo*, the dimorphic element is located about 10,000 base pairs from the *bab1* promoter and over 40,000 base pairs from the *bab2* promoter (Couderc et al.,

2002). However, traditional reporter transgene vectors do not mimic this spacer distance, potentially limiting the Red Light Green Light system. This transgression could be addressed by using a heterologous promoter sequence that is proven to be spatially dependent, so the 2 kb sequence would disrupt the communication between the heterologous promoter and the *bab* enhancer, as mentioned above.

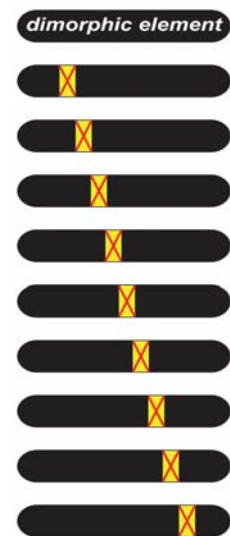
Conclusions on the effects observed from the use of various promoters

For the promoter substitutions, I hypothesized that the endogenous *bab* promoters would be able to interact with the enhancer over a distance, driving the expression of the distal *DsRed2* gene. The results of the promoter comparison (Figure 14) were surprising—the endogenous promoters were less effective than the *hsp70* promoter in driving *DsRed2* expression over the 2 kb distance, despite the fact that they are the targets of this “remote control element” interaction *in vivo*. This tendency may result from a variety of causes, including the efficacy of the *hsp70* promoter that was mentioned in the spacer conclusions sections, and the uncharacterized nature of the “promoter region” for the *bab* gene. The promoter region of the duplicate *bab* promoters was characterized as a ~200 base pair sequence that was directly upstream of the *bab* exons. This “promoter region” was assigned for the sake of this experiment based upon the annotation for the *Drosophila melanogaster* genome project release 5. However, it is uncertain whether the promoter utilized to control *bab* expression by the dimorphic element is the same as that annotated for this genome sequence project. For this reason, the data may not reflect the correct expression of the *bab1p* and *bab2p* interactions with the dimorphic element

enhancer because the “promoter region” may be lacking the specific set of base pairs that is the target of the “remote control element” *in vivo*. Without this unknown sequence, the dimorphic element would lose its ability to contact the endogenous promoters, resulting in the reduced expression of the distal *DsRed2* gene.

One way that we can address this uncertainty is to design promoter regions that encompass more of the region directly upstream of the first exon of the *bab* gene. With more base pairs included, we may unknowingly include promoter proximal sequences that are responsible for this long-distance interaction. Another way that we could address this situation is to explore the dimorphic element in greater detail to determine the specific sequence that is functioning as the “remote control element” (Swanson et al., 2010). By elucidating the exact part of the dimorphic element that is involved in this long-distance interaction, we could then deduce the parameters of the “promoter region” upstream of the *bab* by studying the looping that results from this long-distance interaction. To determine the sequence that is responsible for this potential “remote control element” activity, scanning mutants of the dimorphic element could be made, for which each mutant alters a small segment of the enhancer sequence to see which sequence or sequence is necessary for regulating *bab* gene expression (Figure 15).

Figure 15. Scanning mutation approach to find mutations altering dimorphic element long-distance regulatory capability.



Incorporating a better red fluorescent protein

Though this project did not produce the results we were expecting, it did open many doors to further optimize the Red Light Green Light approach for the mechanistic study of long-distance gene regulation. One avenue to optimize this system is to consider different red fluorescent proteins. This would enhance the efficacy of the system because it would provide a system where the emission spectra for the two fluorescent proteins do not overlap. The system I developed utilizes the DsRed2 gene which makes some protein molecules that fluoresce in the green spectrum. Thus, this signal cannot be easily distinguished from the signal produce by the EGFP gene (Figure 16).

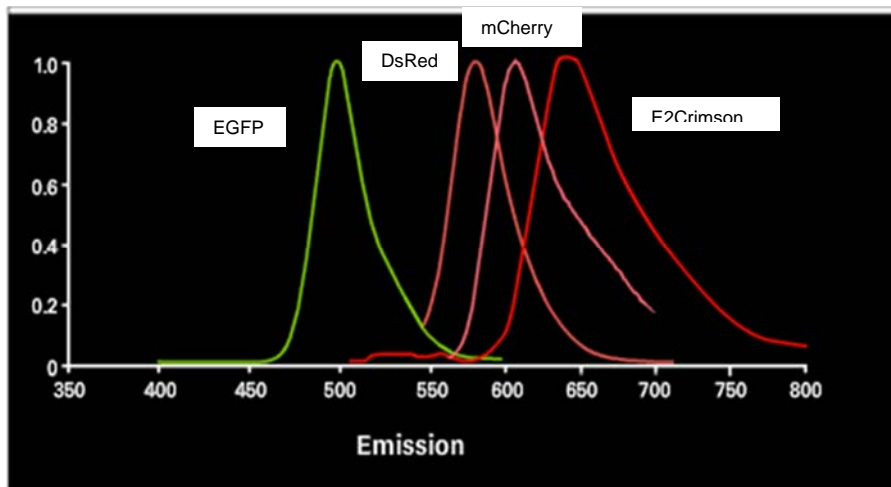


Figure 16:
Emission
spectrum for
EGFP, DsRed,
E2Crimson, and
mCherry

In order to ensure my results were unaffected by this spectra overlap, for each construct tested the fluorescence needed to be calibrated against the original 0 kb vector. This was a tedious process that could be avoided if there was no overlap between the two wavelengths associated with the two fluorescent proteins involved. A red fluorescent protein that does not produce green spectrum fluorescence would eliminate this

complication. Two potentially better reporters are mCherry and E2Crimson, which have red-shifted emission spectra (Figure 16).

Optimization of this system is an integral part of scientific progress and success may provide a method that allows for a better understanding as to how genes are regulated in large genomes and the 3-dimensional space of the nucleus. Because enhancer and promoters are present in all animal genomes, successful preparation and implementation of this vector could shed light on differential gene regulatory pathologies in all species, including humans. Many human diseases have a genetic basis, so understanding the mechanisms behind the differential regulation in a pathological state can have implications in both preventative and pharmacogenomics therapies.

Chapter V: Appendices

A – Phusion PCR

A series of 50ul solutions were created, each containing: 2.5 ul of a sequence specific forward and reverse primers, 10 ul Phusion High Fidelity Buffer, 4 ul 2.5 mM dNTPs, 29.5 ul of Milli-Q, and 0.5 ul Phusion DNA Polymerase Enzyme. Using a thermal cycler (Applied Biosystems, model 2720) samples were run through the following program: One 5 minute cycle at 94°C followed by 10 cycles of 30 seconds at 94°C, 30 seconds at 58°C, and 1 minute at 72°C, followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 1 minute at 72°C, followed by one 7 minute cycle at 72°C, and then samples were held at 10°C until the sample was removed and stored at -20°C for later use.

B – Go-Taq PCR

A series of 20ul solutions were created, each containing: 1 ul of a sequence specific forward and reverse primers, 0.5 ul of template gDNA, 4ul 1X GoTaq Flexi Buffer (Promega, Cat.#M890A), 1.6 ul 2.5 mM dNTPs, 1.2 ul 25 mM MgCl₂, 10.6 ul of Milli-Q, and 1ul GoTaq Polymerase Enzyme (Promega, Cat.#M8291). Using a thermal cycler (Applied Biosystems, model 2720) samples are run through the following program: One 5 minute cycle at 94°C followed by 10 cycles of 30 seconds at 94°C, 30 seconds at 58°C, and 1 minute at 72°C, followed by 25 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 1 minute at 72°C, followed by one 5 minute cycle at 72°C, and then samples were held at 10°C until the sample was removed and stored at -20°C for later use.

C – Gel Electrophoresis

Gel Preparation: Each gel consisted of 100 ml 1X TAE buffer, 1.0 g agarose for 1% gel and 0.7g agarose for 0.7% gel, and 5.0 ul 2 mg/ml ethidium bromide. First, the TAE Buffer and agarose were mixed, and then heated in a microwave on high for approximately 60 seconds. Then, the ethidium bromide solution was added. The gel was then poured into a gel casting mold and allowed to cool before electrophoresis.

Sample Separation: 2.0 ul 4 mg/ml Cresol Red (Sigma-Aldrich, Cat.# 114472-25G) was added to each sample and then the samples were loaded into individual wells of the gel. Electrophoresis runs for 30 minutes at 130 volts. Images of the resulting gels were photographed using a gel documentation system (UVP, BioChemi System).

D – DNA Ligation and Bacterial Transformation

Ligation reactions utilized the T4 DNA ligase (New England BioLabs, Cat.# M0202L) following the manufacturers' protocol. Competent bacterial cells were transformed with ligation reactions to create vectors containing PCR product. The competent cells were retrieved from -80°C storage and thawed on ice for approximately 5-15 minutes. 50 ul of the competent cells were mixed with 2.0 ul of the ligation reaction and allowed to sit on ice for 30 minutes in a disposable snap cap tube. Following ice incubation, bacterial cells were heat shocked for 45 seconds in a water bath at 42°C. Shocked cells were then placed on ice again for 1 minute and then supplemented with 70 ul super optimal broth (SOB) and incubated at 37°C with rotation for 60 minutes. The mixture was then spread on

Luria Broth (LB) petri plates containing Ampicillin at 100 ug/ml. Transformations were allowed to absorb into the agar for 5 minutes and then incubated at 37°C for approximately 18 hours. Successfully transformed bacteria were revealed as Ampicillin resistant colonies.

Luria Broth Plate with Ampicillin: For 1 L of plates, add to 1 L of dH₂O: 10 g Tryptone (Fisher Scientific, Cat.# BP1421-500), 5 g of Yeast Extract (Fisher Scientific, Cat.# BP1422-500), 10 g of sodium chloride (Sigma Aldrich), 20 g microbiology agar (Fisher Scientific, Cat.# BP1423-2). Then adjusted the pH to 7.5 with 10 M NaOH and autoclaved for 30 minutes. When the solution cooled to about 55°C, add 1 mL of Ampicillin (100 mg/mL) was added and media was poured (about 25 mL each) into clean plates, and allowed to cool.

E - Spot Plate Bacterial Transformation

2ul of the DNA of interest is ligated T4 DNA ligase (New England BioLabs, Cat.# M0202L) following the manufacturers' protocol. Next, the bacterial transformation procedure following a normal ligation was utilized with the vector of interest. The transformed bacteria were plated on LB-Amp plates and incubated at 37°C for about 18 hours.

F – QIAGEN Mini Prep Procedure

Cultures from a transformation were incubated with LB medium-Ampicillin at 37°C for 18 hours in snap-cap tubes with rotation. The mini-prep cultures were transferred from snap-tubes to 2ml flat tubes. The culture was then centrifuged at 13,000xg for 10 minutes, and the supernatant was discarded, until the entire culture had been spun down into the 2ml tube. Next, the pellet was resuspended in 250ul of Buffer P1 (stored at 4°C), and set to rest for 3 minutes. The sample was then vortexed for 30 second to fully resuspend the pellet, 250ul of Buffer P2 was added, and the sample was set to rest for 5 minutes. Then, 350ul of Buffer N3 was added and the tube was inverted several times to mix the solution. The samples were then placed in the -20°C freezer for 10 minutes, followed by a 10 minute centrifuge to isolate the supernatant. The supernatant was then transferred to spin columns, and centrifuged for 2 minutes. The DNA was retained by the column, while the excess supernatant was discarded, and 750ul of Wash Buffer PE was added and centrifuged through the spin column twice. The spin column was then transferred to 1.5ml tube, and 60ul of Elution Buffer (EB) was added, and the sample was set to rest for 1 minute, while the DNA was eluted from the column. The tube was then centrifuged for 2 minutes, and the spin columns were discarded. The completed mini-prep samples are stored at 4°C in the short term and -20°C in the long term.

G - Ethanol Precipitation

For a 50ul digest, add 5ul of 3M NaOAc, 140ul of 100% ethanol, and 1ul of glycogen. This solution was placed in a -20°C freezer for 30 minutes, then centrifuged at 13000xg

for 30 minutes. The centrifuge isolated the precipitated DNA as a pellet at the base of the tube. The supernatant was aspirated off. 1ml of 70% ethanol was then added to the DNA solution for further purification. This solution was centrifuged for 15 minutes, and then the supernatant was removed by aspiration. The purified DNA sample was then resuspended from a pellet into 50ul of EB Buffer from the QIAgen Purification Kit.

H – PCR Purification

5 volumes of Qiagen Buffer PB are added to each reaction in an appropriately sized microfuge tube and mixed. Purple columns for each of the purifications are labeled and the sample mixed with Buffer PB is added to the top of the purple column. The columns are then spun in the microfuge for 2 minutes at high speed, and the flow through from the collection tube is dumped out. Next, 750 ul of Qiagen Buffer PE is added to the top of each purple column, and the columns are spun in the microfuge for 2 minutes at high speed (~10000 revolution per minute). The flow through is dumped out of the collection tube and the columns are placed back in the tube, and spun again for 2 minutes. 1.5 ml microfuge tubes are labeled appropriately and the purple columns are placed into this fresh tube. Finally, 35 ul of Qiagen Buffer EB is added to the top of each column. The columns are left to sit while the buffer interacts with the immobilized nucleic acids for one minute and then the column and tube are spun in a centrifuge for 2 minutes at high speed. The purple columns are thrown away as the DNA of interest is now in the buffer EB solution in the 1.5 ml tube, and the tubes are stored at -20°C.

I – Purifying Small DNA fragments with the Qiagen QG Kit

Using a razor blade and a UV light box, the DNA of interest is cut and put into a separate 2.0 ml microfuge tube. The agarose slab is then weighed and 3 volumes of Qiagen QG Buffer is added to each tube (1 volume equals 100 ul per .1 g of agarose). The samples are then incubated in a 50-65°C water bath until the agarose dissolves, mixing by flicking every 5 minutes. After the agarose has melted, 740 ul of the melted agarose solution is added to the top of a purple Qiagen column and the solution is spun through the column for 2 minutes at high speed in a microfuge. Then the column flow through is dumped out of the collection tube, and this step is repeated until all of the dissolved agarose solution has passed through the purple column.

After, the DNA is washed by adding 750 ul of Qiagen Buffer PE to the column and spinning this wash solution through the column using the microfuge at high speed. The wash buffer is then dumped out of collection tube, and the purple column is placed back in collection tube and spun one more time. Finally, the purple column is transferred to a fresh 1.5 ml microfuge tube that is labeled and dated appropriately, and 35-60 ul of Qiagen Buffer EB is added to each column, allowed to sit for 1 minute, and then spun in microfuge for 2 minutes at high speed. The purple columns are then discarded and the purified DNA is stored at -20°C.

J – Purifying Large DNA Fragments (>6 kb) Using the Qiagen QX1 Kit

Using a razor blade and the Williams lab's UV light box the desired DNA fragments are cut out, placed in separate 15 ml snap cap tubes, and weighed on the scale. 3 volumes of

QX1 buffer are added to the snap cap tube with agarose slab (1 volume equals 100 ul per .1 g of agarose) with 2 volumes of milli-Q and 35 ul of QIAEX II Suspension (slime). The snap tube is closed tight and vortexed for a few seconds to mix the components. Next, the sample is incubated at 65°C until the agarose dissolves, while the sample is briefly vortex every 5 minutes. After the DNA has melted, 1.8 ml of the dissolved gel solution is added to a 2 ml tube and the sample is spun in the microfuge for 2 minutes at high speed. Then, the supernatant is dumped into the trash or sink and the process is repeated until nearly all of the solution from the snap cap tube has been spun down in the 2 ml microfuge tube.

After dumping off the last of the supernatant from the 2ml tube, 0.75 ml of fresh QX1 buffer is added to the tube, and vortexed until the pellet has been resuspended. Then the samples are microfuged for 2 minutes at high speed in the microfuge and dump off the supernatant in the trash. The isolated DNA sample is then washed as 0.5 ml of Qiagen Buffer PE is added to each sample and briefly vortexed to resuspend the pellet. Then the samples are spun in the microfuge for 2 minutes at high speed, and the supernatant is dumped off. Then this wash step is repeated. Next, the supernatant is aspirated from the microfuge tube, trying to remove most of the liquid and placed in a heat block set to 50°C with lids open and the samples are dried for 5-10 minutes.

Finally, 50-75 ul Qiagen Buffer EB is added to the microfuge tubes and the tubes are vortex to resuspend the pellet. To elute the DNA, the tubes are floated in a water bath or incubator set to 65°C for about 10 – 30 minutes, flicking every 5 minutes. To isolate the eluted DNA, the tubes are spun in the microfuge for 3 minutes at high speed, and

supernatant is removed by pipetting as it is transferred to a fresh 1.5 ml microfuge tube.

The isolated DNA is stored at -20°C.

K – UV spectroscopy to determine DNA concentrations

The nanospec machine was used to quantify the concentration of mini/midi prepped DNA. This machine is able to detect the concentration of the DNA suspended in elution buffer following a mini or midi prep procedure. The machine is first calibrated with standard elution buffer provided in the Mini Prep Kit from New England BioLabs. As light passed through the column in the machine, the concentration is quantified based on the amount of light that can pass through the sample. This quantification is important for sequencing the DNA because the ideal concentration to sequence DNA is 70nm/ul. This standard is reached using the nanospec to determine the original concentration, and the solution is diluted with milli-Q water to reach the 70nm/ul concentration.

L – Fly Sugar Food Preparation

The food contains per liter: 1000 ml distilled H₂O, 48.53g yellow corn meal (Meijer brand), 10.27g granulated sugar (Meijer brand), 7.37g agar (Colony Processing, CAD 2188486), 25.67g brewer's yeast (MP Biomedicals Cat.# 903312), and 2.38g anti-fungal Methyl-4-hydroxybenzoate (Sigma-Aldrich, CAD 99-76-3). Ingredients are mixed into boiling water, poured into fly vials, and cooled overnight at room temperature while covered with cheese cloth. Food can be stored for several weeks at 4°C while covered by a garbage bag.

M – EcoRI-DsRed2-AscI Sequence

The following sequence was designed by Eric Camino and synthesized by GenScript Inc. that contains the DsRed.T4-NLS coding sequence (Barolo, Castro, Posakony, 2004) that had an internal SbfI site ablated by a synonymous mutation. The DsRed.T4-NLS gene has the *hsp70p* located just 5' and flanked by StuI and AscI restriction enzyme sites. The EcoRI site and an FseI site flank the DsRed2 gene 3' untranslated region.

```
GAATTCATTAACGCTTACAATTTACGCCTTAAGATACATTGATGAGTTTGGG
CAAACCACAACACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTG
ATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAACAAGTTAACAAC
ACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTT
TTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGCTGATTATGATCTAGAG
TCGCGGCCCGCCGCTGTACAAGTAGCGTCTTCGTTCACTGCTGCGACTTCGGC
TCCGTTGACGGCGTTTTTCGTTCACTGCTGCGACTTCGGCTGCGATTGCGGCTC
CTAGAGCGCTGGCGATGCCTTCTGTGCCTGCTCTTGTACAGGAACAGGTGGTG
GCGGCCCTCGGCGCGCTCGTACTGCTCCACGATGGTGTAGTCCTCGTTGTGGG
AGGTGATGTCCAGCTTGGAGTCCACGTAGTAGTAGCCGGGCAGCTGCACGGG
CTTCTTGGCCATGTAGATGGACTTGAAGTCCACCAGGTAGTGGCCGCCGTCCT
TCAGCTTCAGGGCCTTGTGGATCTCGCCCTTCAGCACGCCGTCGCGGGGGTAC
AGGCGCTCGGTGGAGGGCTCCCAGCCCATAGTCTTCTTCTGCATTACGGGGC
CGTCGGAGGGGAAGTTCACGCCGATGAACTTCACCTTGTAGATGAAGCAGCC
GTCCTGCAGTGAGGAGTCCTGGGTACGGTCACCACGCCGCCGTCCTCGAAG
```

TTCATCACGCGCTCCCACTTGAAGCCCTCGGGGAAGGACAGCTTCTTGTAGTC
 GGGGATGTGCGCGGGGTGCTTCACGTACACCTTGGAGCCGTACTGGAAGTGG
 GGGGACAGGATGTCCCAGGCGAAGGGCAGGGGGCCGCCCTTGGTCACCTTCA
 GCTTGGCGGTCTGGGTGCCCTCGTAGGGGCGGCCCTCGCCCTCGCCCTCGATC
 TCGAACTCGTGGCCGTTACGGAGCCCTCCATGCGCACCTTGAAGCGCATGA
 ACTCCTTGATGACGTCCTCGGAGGAGGCCATGGTGGCGACGGGTGAAGGGCG
 AATTAAACAGGCCTCTCTCGACTCTAGCGCGTACCCTAGAGGATCAGCTTGG
 CTGCAGATTGTTTAGCTTGTTTACGCTGCGCTTGTTTATTTGCTTAGCTTTCGCT
 TAGCGACGTGTTCACTTTGCTTGTTTGAATTGAATTGTCGCTCCGTAGACGAA
 GCGCCTCTATTTATACTCCGGCGCTCGGCGCGCC

N – AscI-Dimorphic Element, 1.0 kb spacer-AgeI Sequence

The following sequence flanked by AscI and AgeI restriction endonuclease sites was designed by Eric Camino and synthesized by GenScript Inc., which contains the dimorphic element (Williams et al. 2008) with an adjacent 1.0 spacer sequence, followed by the *hsp70p*. The dimorphic element was flanked by the AscI site and SbfI and NheI sites on the other side. The 1.0 kb spacer was made by taking 1.0 kb of *babI* intron sequence and altering every other base pair by a non-complementary transversions. The spacer was flanked by SbfI and BglII restriction endonuclease sites. The *hsp70p* is flanked by a BamHI site on the 5' side and SacI and XhoI sites on the 3' side.

GGCGCGCCCCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATA
 TAAAAAAGCTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAA

AAGTTGCAGGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAA
CAGCAGCAACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCA
ACAATGTTGCTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAA
TTATATGCAACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAA
AGTGGAGTGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATT
GGCGCAAGTTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACA
ATTTTACCATTTTCAGCCACAACCTTTTCGCACTGCTCCCCCCTTTCCCAGCACA
ACAATGTTGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTTATATCACTTA
CTTTACTTAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATT
TGCTCCAGCAGTGGGCTGCATTTTACGACCCTCAAACCCGATCCAAATGGA
AAATATGAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACAC
ACGCTAGCCCTGCAGGACCACTAGCAACTGGCTCGACGATCATCGGCCCA
CCGAACCGATGAATGTCAGATTAGGACAGAGTAATAGAACTTTTGTGCGGG
CAATGAGTTACAGAGACTGAGGCATCCACGACTCCTGATGCCGAGATAGAGC
GTAAGAGGGAGAGTTCCATCTCCTATGACGCTAACTCTGTCAGAGTGTGTGT
GTCAGTGTAACCTCTAATCCTGTAACACTGTCACAGAGAGTGTGAGGAACTAA
GACTTATGAACTATAATACATATACACAGAAAGACACACTACCACTCAGGCA
GCGAGCTAGTGTCGACACCGCTCTGTGCGGGGAATAACAAGTACTTAGGTTC
ACGACTGGCATCCCCGTCGTGCAACCCCAAACGGCCCCACAGGATACCTCG
CAAGCACTAACCAGAATGACCCCGTACTGTAACCTAAACGCGTTGGCCCAGCA
TCTTGCAGTTTCATTCGCACAGCGCAACAACGCAGCTGGGTAGCGGCCATTAT
ACGTGTTCCGACAGAATCTTTTGCGAACAGCGCGGAGTGCATACATATCTGA
CTATGCGCGAACAACCTTGCAAGGGTTCCTTTGCGAAAGGGCAGGATTTAAC

TAAGTACTGTCCAGGTCAAATATTGAGTTTGTGGGTGTATTGCAAAATATCTC
GGAGCCAGGATCGCAGAAGTACTCTCGGGGCCTTAAAGCCCCTATGAGCACT
TCAGACTGAATACTGCTCCGACTACGCTGGCATAGCTCAGCACGTCTCTCTAC
TAGGGCAGCCACAAACAACCTATGAACACGCTAGTAGACTTACTTGTATATAT
AGAGATAAGTACGAATCTCGGAGCCCGAGAATTTAAATCTCGGAATCCGAAT
ACGCGCGATACCCAGAGAGATGCGCGCGCGCGCGCGCGCGGACGCGCGTGT
GGGATACTGGTACTGGAATACTGGCATATTATGCGCGTCAATTGAAGAGGGA
ATCAATGATACACCTAGTAAAACCCTCTTACTCAGTTACACCCACATTCCTAC
ATTGAGCGTGACGTACGCTTACAGAAACATAGACAAGGGCACTGTCATTGGA
TGATAGTAAATAAAAGATCTGGATCCGAGCGCCGGAGTATAAATAGAGGCGC
TTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCT
AAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCT
GCAGCCAAGCTGATCCTCTAGGGTACGCGCTAGAGTCGAGAGCTCGAGAGCC
CTGTTTAATTCGCCCTTCACCGGT

O – 3,103 base pair StuI-AgeI fragment with 2.0 kb spacer

The following sequence flanked by StuI and AgeI restriction endonuclease sites was designed by Eric Camino and synthesized by GenScript Inc., which contains a 2.0 kb spacer sequence and the dimorphic element. In detail, an *hsp70p* is present and flanked by StuI and BamHI restriction sites. This promoter is followed by a 2.0 kb spacer sequence that was derived from 2.0 kb of *babI* intron sequence that was mutated at every other base pair by non-complementary transversions. The spacer sequence was followed

by a BglII site, then the dimorphic element flanked by AscI and SacII sites on one side and NheI and SbfI sites on the other side. Finally, this fragment contained a second *hsp70p* flanked by XhoI and AgeI restriction endonuclease sites. The sequence is:

```
AGGCCTCTCGACTCTAGCGCGTACCCTAGAGGATCAGCTTGGCTGCAGATTGT
TTAGCTTGTTTCAGCTGCGCTTGTTTATTTGCTTAGCTTTCGCTTAGCGACGTGT
TCACTTTGCTTGTTTGAATTGAATTGTCGCTCCGTAGACGAAGCGCCTCTATTT
ATACTCCGGCGCTCGGATCCAGTTACAGAGACTGAGGCATCCACGACTCCTG
ATGCCGAGATAGAGCGTAAGAGGGAGAGTTCCATCTCCTATGACGCTAACTC
TGTCAGAGTGTGTGTGTCAGTGTA ACTCTAATCCTGTAACACTGTCACAGAGA
GTGTGAGGAACTAAGCATTAGTAACTATAATACATATACACAGAAAGACACA
CTACCACTCAGGCAGCGAGCTAGTGTCGACACCGCTCTGTGCGGGGAATAAC
AAGTACTTAGGTTACGACTGGCATCCCCGTCGTGCAACCCCAAACGGCCC
CACAGGATACCTCTCTAGCACTAACCAGAATGACCCCGTACTGTA ACTAAAC
GCGTTGGCCCAGCATCTTGCA GTTTCATTTCGCACAGCGCAACAACGCAGCTG
GGTAGCGGCCATTATACGTGTTCCGACAGAATCTTTTGCGAACAGCGCGGAG
TGCATACATATCTGACTATGCGCGAACAACTTGCAAGGGTACCTTTGCGAA
AGGGCAGGATTTAACTAAGTACTGTCCAGGTCAAATATTGAGTTTGTGGGTG
TATTGCAAAATATCTCGGAGCCAGGATCGCAGAAGTACTCTCGGGGCCTTAA
AGCCCCTATGAGCACTTCAGACTGAATACTGCTCCGACTACGCTGGCATAGC
TCAGCACGTCTCTCTACTAGGGCAGCCACAAACA ACTATGAACACGCTAGTA
GACTTACTTGTATATATAGAGATAAGTACGAATCTCGGAGCCCGAGAATTTA
AATCTCGGAATCCGAATACGCGCGATACCCAGAGAGATGCGCGCGCGCGCGC
GCGCGGACGCGCGTGTGGGATACTGGTACTGGAATACTGGCATATTATGCGC
```

GTCAATTGAAGAGGGAATCAATGATACACCTAGTAAAACCTCTTACTCAGT
TACACCCACATTTCCTACATTGAGCGTGACGTACGCTTACAGAAACATAGACA
AGGGCACTGTCATTGGATGATAGTAAATAAACTGGCACTCTCAGTCTTACGA
ACATCAGAAGTCGTAATCTCGCTCTATGCCTCTTTCTCTGGAACGAGAAGCGT
CATAGCCAGAGTGACTCTGTGTGTGTGACTGTGCCAGAGCCGAAGTGCCACA
GTGACACTCTCTGTGTCTTCCAGCCTCAGGCTGCCAGCGCCGCACGCGCACAC
TCCCTCACACAGCAACAGACTTACTATCTAGCTGTGATCACAATAGAGTGTAT
TTTCCGCCAACTGCAGGCTTGGACATCAGTTACGAAAAATGATGTACCAAAA
CCCATTA AAAACACTTCGCAAGATAGCTACAGCCAACTCCGTCAAAATGCAGT
GCCAGCCCATATGGTTAAACTACGAGGTACTGGGACGGATACACTATACCAC
CATTCTAGTTTGCTATTAACGGCGCATGTGGAATCACTCCGAGGGGTATCCAC
TATATTCTGTACGCACGCGAGTCAGCGTATATCCACCCAGGTACCTTTGCAAG
GGTATCCCTTTACTTCGGGCCAGCCTGCAGTGA ACTTGACCAGCGGTCTGGGT
GTTTGTGCGGTACCCCGCGAGATTCTAACTTCGATACTCCTGCAGAGATTTTA
AGGCCCTAAAAGCGTCTACAGGACTCAGTCCGCAGTAGAATCAGCATAGTTA
CGCTAGACTACATGAGAGAGCAGCTTTACTAACACCCACCAGCGTCCACATA
GCTGCTCAGGCAGGTGCGCGCGCTCTCGCCTGCATCCGAGATTCTAAATCTCC
GGGCCCCGAGATTCCGAATCCGCATATATCGCAA ACTCTCTCGTATATATATAT
ATATATATTCATATATGTGTTTCGCAGTTGCAGTTCCGCAGTTACGCGGCGTA
TATGTCCGGTCCTCTTTCCGACCGTCGCACAAGCTGCCCCAAAGAGGCAGAC
TGGCACAAACACGGAAGCACGGTCTATGTCATGCATAGGCACTCCCACGCTC
ACCTTTACAGTGACGGTTCGTGCTGCCCCGCCAGATCTGGCGCGCCTCGCCT
CCGCGGCTCTTTCTCTTTGCCATTTTAACTTTTATTACTCTTAATATAAAAAAG

CTGGCTAGATGCGGGCCAGCTGTAAAAATGCACGCGGTCATAAAAAGTTGCA
GGAGGCATGTTGCCAGTTGCCTGCAACCGGCAACATTCGCAGAACAGCAGCA
ACATCGTAAAATAACTTCTTGCTCTGCGGTCTGAGTTTGGCCGCAACAATGTT
GCTGCATTTATTCGTATTATTATTACATTTTAATGAATAATTCTAATTATATGC
AACTTGAATAAGCCCGCCGATGCCAATAAAAAGCGGCGTGGCAAAGTGGAG
TGGACTGGGTTTGTGTGGCGCCCCTGCTAGTGGCACATAAAAATTGGCGCAA
GTTAATTGTGGTAGTTATTTGCTGTTTTGCCATTTGGTCATTTTACAATTTTAC
CATTTCAGCCACAACCTTTTCGCACTGCTCCCCCCTTTCCCAGCACACAATG
TTGCGGCATTCTCGCACTTTACGAGGCGTTTTTTTTTTATATCACTTACTTTAC
TTAGTTGATTAAGGGCGTGGCCGATGGGCCAGATACATGCTTAGATTTGCTCC
AGCAGTGGGCTGCATTTTACGACCCTCAAACCCGATCCAAATGGAAAATAT
GAAAATACGGCTAATCCGCTTATGAGCACAACAAATTGGTTCACACACGCTA
GCCAAGGGCCTGCAGGCTCGAGGAGCGCCGGAGTATAAATAGAGGCGCTTC
GTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAG
CGAAAGCTAAGCAAATAACAAGCGCAGCTGAACAAGCTAAACAATCTGCA
GCCAAGCTGATCCTCTAGGGTACGCGCTAGAGTCGAGAGCCCTGTTTAAACG
ATCCACCGGT

Chapter VI: References

- Barolo, S., Castro, B., & Posakony, J. W. (2004). New drosophila transgenic reporters: Insulated P-element vectors expressing fast-maturing RFP. *Biotechniques*, 36(3), 436-40, 442.
- Brown, T. (2002). Genome anatomies . *Genomes* (2nd Edition ed., pp. Chapter 2) Oxford: Wiley-Liss.
- Couderc, J. L., Godt, D., Zollman, S., Chen, J., Li, M., Tjong, S., Laski, F. A. (2002). The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in drosophila. *Development (Cambridge, England)*, 129(10), 2419-2433.
- Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic drosophila by using the site-specific integrase from phage phiC31. *Genetics*, 166(4), 1775-1782.
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945.
- IUM, T. (2012). ENCODE project writes eulogy for junk DNA.
- Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., de Graaff, E. (2003). A long-range shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14), 1725-1735.
- Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E., & Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nature Genetics*, 40(4), 476-483.
- Rebeiz, M., Castro, B., Liu, F., Yue, F., & Posakony, J. W. (2012). Ancestral and conserved cis-regulatory architectures in developmental control genes. *Developmental biology*, 362(2), 282-294
- Rebeiz, M., & Williams, T. M. (2011). Experimental Approaches to Evaluate the Contributions of Candidate Cis-regulatory Mutations to Phenotypic Evolution. In *Molecular Methods for Evolutionary Genetics* (pp. 351-375). Humana Press.

- Rogers, W. A., Williams, T. M. (2011). Quantitative comparison of cis-regulatory element (CRE) activities in transgenic *Drosophila melanogaster*. *Journal of Visualized Experiments*, (58)
- Rogers, W. A., Salomone, J. R., Tacy, D. J., Camino, E. M., Davis, K. A., Rebeiz, M., & Williams, T. M. (2013). Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. *PLoS Genetics*, 9(8), e1003740.
- Salomone, J. R., Rogers, W. A., Rebeiz, M., & Williams, T. M. (2013). The evolution of bab paralog expression and abdominal pigmentation among *Drosophila* fruit fly species. *Evolution & Development*, 15(6), 442-457.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning*. Cold spring harbor laboratory, press. New York.
- Swanson, C. I., Evans, N. C., & Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a notch-and EGFR-regulated eye enhancer. *Developmental Cell*, 18(3), 359-370.
- Visel, A., Bristow, J., & Pennacchio, L. A. (2007). Enhancer identification through comparative genomics. *Seminars in Cell & Developmental Biology*, 18(1) 140-152.
- Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell*, 134(4), 610-623.
- Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., & Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, 389(1), 52-65.